

On the Impact of Diagram Layout: How Are Models Actually Read?

Harald Störrle¹, Nick Baltsen¹, Henrik Christoffersen¹, Anja M. Maier²

¹Dept. of Applied Mathematics and Computer Science

²Dept. of Management Engineering, Technical University of Denmark

Abstract. This poster presents the latest results from a very large eye tracking study (n=29) that explores how modelers read UML diagrams. We find that various factors like layout quality, modeler experience, and diagram type lead to significant differences in diagram reading strategies. We derive elements of a theory of diagram reading behavior from our findings. This paper presents only late breaking results: all findings presented, theories constructed, and conclusions drawn are of a preliminary nature. This paper does not present the amount and degree of evidence that would allow us to consider the contents as being scientifically validated, yet.

1 Introduction

Practical experience suggests that usage and understanding of UML diagrams is greatly affected by the quality of their layout: it is a lot harder to spot errors and weaknesses in a cluttered diagram, and it is a lot easier to understand the—quite literally—overall picture of a nicely laid out diagram. Previous research has identified what constitutes good and bad practices of layout (see e.g. [1]), but has failed to provide conclusive evidence in support of this hypothesis. In fact,

“We set up an experiment to study the impact of [layout] rules on understandability [...] operationalized in terms of [...] faults and [...] time. We could not identify a statistically significant relation between [them].”[1]

On the other hand, our own previous work provided substantial evidence to this effect [4, 5]. We could demonstrate, for instance, that error rate, time used, and subjectively assessed cognitive load increased for UML diagrams with bad layout, and that this effect is likely independent of the diagram type (we studied the five most commonly used diagram types of UML).

Since there is such a stark contrast between our results and the previous literature, we deemed it necessary to add to the validity of our study by a replication of the results. The replication varies the previous studies in three aspects.

- First, obviously, different participants were recruited for the present study. They are of a comparable population, though, so this amounts to a minor variation point only.

- Second, the study is designed and conducted by two graduate students (the second and third authors), so as to reduce impact and bias through the person of the experimenter of the previous studies (the first author). Again, this was expected to be of secondary importance due to the previous study setup that had reduced personal interactions almost entirely.
- Third, we use a different method (eye tracking) with objective, physiological measurements of cognitive load, thus addressing critique that subjective assessments of cognitive load are not sufficiently reliable. We also used the measurements previously employed (score, duration, subjective assessment) to establish a point of reference.

Next we were curious whether or not we would be able to observe any identifiable behavior patterns, that amount to reading strategies. If such reading strategies do exist, would there be differences under the various treatments, e.g., would novices exhibit different reading strategies than experts, say? Ideally, this should lead to elements of a theory of how UML models (and similar artifacts) are processed by human modelers, and what we can do to support this process. Since these are subtle questions, and since we have not used eye tracking as a research method before, we have developed and refined our study design and experimental procedure in a pilot study, see [3] (n=4). Here, we report initial results from the main study following on the pilot study (n=29, one excluded due to technical errors in the measurement process).

2 Research Objective 1: Replication of previous results

By and large, the present study produced the same results as previous studies: the impact of layout quality on modeler performance could be replicated, though it showed more in the objective measures (score, time) than in subjective effort assessment – in our earlier studies, we found a stronger effect in cognitive load than in objective achievement. The differences are not so big as to be inexplicable, though, which leads us to attribute it to the different experimental setting, and the different populations. As in earlier experiments, the effect appears independent of diagram type, but is influenced by the expertise level, and the size of diagrams.

Previously, we had only used subjective assessments of cognitive load which have been criticized as less reliable than physiological indicators like heart rate or skin conductivity. So we instrumented our stimuli accordingly and measured the pupil dilation, fixation duration, and blink rate associated to diagram elements and easily localizable layout problems (e.g., line crossings, line bends). We could clearly show that subjective assessments correlate tightly to physiological measures, which is in line with existing literature on the subject [2]. We also found, that there is small, yet statistically significant evidence that one objective performance indicator (score) *increases* with decreasing layout quality. Obviously, this is counter-intuitive and in contrast to all previous findings. So, while we lack an explanation for this observation so far, we tend to attribute this to a mistake in the experimental procedure.

3 Research Objective 2: Existence and identification of behavioral patterns

We could clearly identify behavioral patterns in the subjects reading diagrams. In order to study these patterns, we recorded and analyzed the starting point and progress of participants' scan paths. We found three different behavioral patterns in them.

- **Branch Following From Anchor (BFFA):** An anchor would be established either at the top left corner, the largest element, or the center of the diagram. Starting from there, the graph structure of the diagram would be followed in a systematic way.
- **Left/Right Top/Bottom (LRTB):** starting at the top left corner, proceed downwards and to the right irrespective of the direction of arrows in the diagram, i.e., as if reading a text in the normal reading direction for western languages.
- **Random Walk (RW):** Starting at any point, continue to any other point and so on, at no discernible pattern.

LRTB coincides with BFFA for diagrams of a diagram type with natural direction (e.g., Activity Diagrams), when the actual diagram layout agrees with the reading direction. In the literature, this has been described as “diagram flow”, a notion that could be formalized as the dominant direction of directed elements in a diagram. This definition does not cover other kinds of flow like circular or radial, but our study did not consider examples of these layout types.

Interestingly, all participants (i.e., both experts and novices) followed the BFFA or LRTB strategies for diagrams with good layout, irrespective of diagram type. For diagrams with bad layout, however, novices tended to use the RW strategy while experts continued to use the BFFA/LRTB strategies.

4 Research Objective 3: Analysis of underlying cognitive processes

The stability of behavioral patterns we have observed were relatively stable over different individuals, and even more so for advanced modelers. This might be interpreted to show that while behavior exhibits individual variation in less skilled modelers, increasing experience results in a converge of behavior. If this is so, it is likely that one and the same cognitive process is effective *universally*, in all modelers—but what can we know about this process?

In a recent study [6], we could show that an important factor for decreasing modeler performance (and increasing difficulty) is the size of the diagram, as measured by the number of model elements. Our present study showed, however, that diagram flaws like line crossings exhibit similar cognitive load profiles as shapes (i.e., diagram elements representing classes, use cases, and so on), as measured by pupil dilation, fixation duration, and blink rate. That means, that

increasing (or decreasing) the number of diagram elements has the same effect on cognitive processing than increasing (or decreasing) the number of layout flaws. This seems to suggest that diagram size and layout quality are just two sides of a coin, and the same cognitive process is affected by both.

The literature on layout quality, however, frequently discusses two distinct layers: one concerned with low-level layout flaws like line crossings, and one concerned with higher-level aspects like flow (this latter level is sometimes called diagram architecture). And indeed, as we have seen above, there are clearly discernible reading patterns that obviously are independent of diagram size, but do depend on the overall layout structure. So, there is some indication that diagram reading might involve at least two different cognitive processes, but it is not quite clear whether and how they are connected.

For diagram types with a “natural” flow like Activity Diagrams, it is fairly obvious to assume that following the reading direction (strategy BR) leads to good layouts. So we might be tempted to consider “layout flow” as a layout quality measure. But what does flow mean for diagrams with no natural directionality? Also, our sample did not contain diagrams with radial or circular layout where different flow might be observed. Finally, it might be that flow is the wrong notion to begin with, and we should rather consider flow as a special kind of symmetry, and explore this notion instead. This will require experiments on specifically created layouts that highlight one or the other aspect, and to see whether these layout styles give rise to different behavior.

5 Interpretation

We conduct a study on the impact of layout quality, expertise level, and diagram type on the understanding of UML diagrams. Like in earlier studies, we find that the diagram understanding *outcome* (measured in time, errors, and load) is affected by layout quality, even though in the present experiment, we used eye tracking instead of the subjective assessments of cognitive load we had used before. We conclude that our initial study and our replication corroborate each other, underlining the validity of the evidence presented.

But it is not just the outcome that is affected by layout quality, we also observe changes to the diagram reading behavior (measured as start location and subsequent evolution of the scan path). Even though the results presented here are of preliminary nature they clearly indicate that there are several distinct reading strategies. Surprisingly, we find the same reading strategies for different diagram types: one would expect that diagram types with a natural directionality might be read in a different way than diagrams without a “built-in” direction. More research is needed to explore this question further, possibly the experimental setup of our study (in particular, the respective layouts) limit the insight of our study.

We have replaced the global and undifferentiated assessment of cognitive load used in earlier experiments by three concurrent, objective measurements with high temporal and spatial resolution: pupil dilation, fixation duration, and blink

rate. This allows us to identify the cognitive load of *individual diagram elements*, and we observe that shapes (i.e., diagram elements representing classes, use cases, actions, and so on) provoke a similar cognitive load profile than layout flaw such as line crossings and line bends. In other words: increasing the number of diagram elements has the same effect on cognitive processing than increasing the number of layout flaws. This explains in a very elegant way our previous finding that diagram size seems to be more important than layout quality: the number of diagram flaws can (often) be reduced, the diagram size (measured in the number of elements, [6]) cannot.

6 Summary and Future Work

We replicate earlier experiments on the impact of layout quality (and other factors) on UML diagram understanding by using eye tracking instead of subjective assessment. We observe characteristic behavior patterns and study several influence factors. We observe a great uniformity in the reading strategies applied (at least for experienced modelers) to the degree we are almost certain to observe a universal cognitive process. If it is indeed universal, we might be able to find a physiological substrate to the observed behavior. Obviously, our present study is not suited to explore this thread further, and we have to defer that to future research employing different machinery, in particular brain imaging.

References

1. Holger Eichelberger and K. Schmid. Guidelines on the aesthetic quality of UML class diagrams. *Information and Software Technology*, 51(12):1686–1698, 2009.
2. Daniel Gopher and Rolf Braune. On the Psychophysics of Workload: Why Bother with Subjective Measures? *Human Factors*, 26(5):519–532, 1984.
3. Maier, Anja M. and Baltsen, Nick and Christoffersen, Henrik and Strrle, Harald. Towards Diagram Understanding: A Pilot-Study Measuring Cognitive Workload Through Eye-Tracking. In *Proc. Intl. Conf. Human Behavior in Design*, 2014.
4. Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams. In *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC'11)*, pages 135–142. IEEE Computer Society, 2011.
5. Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams: Diagram Type and Expertise. In Gennaro Costagliola, Andrew Ko, Allen Cypher, Jeffrey Nichols, Christopher Scaffidi, Caitlin Kelleher, and Brad Myers, editors, *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC'12)*, pages 195–202. IEEE Computer Society, 2012.
6. Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams: Size Matters. In Jürgen Dingel, Wolfram Schulte, Isidro Ramos, and Emilio Abrahao, Silviaand Insfran, editors, *Proc. 17th Intl. Conf. Model Driven Engineering Languages and Systems (MoDELS'14)*, number 8767 in LNCS, pages 518–534. Springer Verlag, 2014.